

2025 新教材技术高考第一轮

专题三 数据处理与应用

考点过关练

考点一 常用表格数据的处理

1.(2022 诸暨海亮高中期中,16)为了响应全民健身计划,某高中对学生进行了体质健康测评,并用 Excel 软件进行数据处理,如图 a 所示。

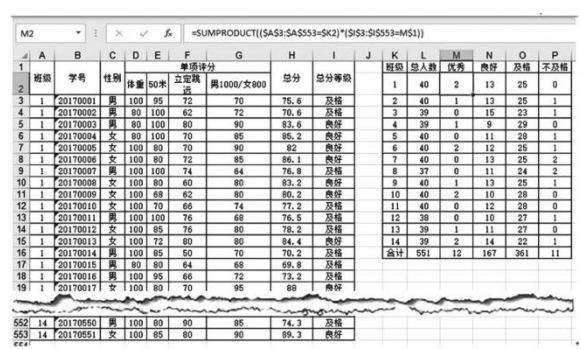


图 a

请回答下列问题:

(1)区域 M2:P15 的数据是通过公式计算得到的,在 M2 单元格中输入公式后,再使用自动填充功能完成该区域的计算,则 N3 单元格中的公式是“=SUMPRODUCT((____)*(\$I\$3:\$I\$553=N\$1))”。

(提示:M2 单元格输入公式=SUMPRODUCT((A1:A100=“2”)*(I1:I100=“优秀”)),表示同时满足 A1:A100 是 2 班和 I1:I100 是优秀这两个条件的情况数量,也就是进行条件计数。)

(2)根据图 a 中的数据制作的图表如图 b 所示。创建该图表的数据区域是_____。

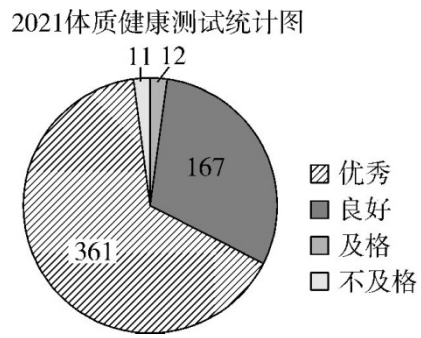


图 b

(3)对图 a 工作表进行相关数据分析,下列说法正确的是_____(多选,填字母)。

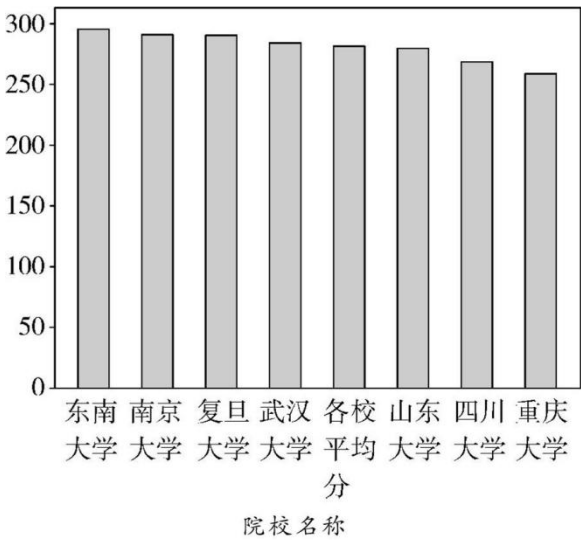
- A.以“总分等级”为依据进行升序排序,选择的区域是 A3:I553
- B.选择“A3:I553”按照“总分等级”升序排序后,图 b 的图表不会发生改变
- C.现需要选出总分等级为‘不及格’的学生,可选择区域 A3:I553,以“总分等级”为‘不及格’进行筛选并保留结果
- D.为了显示男生立定跳远得分最高的 3 位学生数据,可先对“性别”为男”进行筛选,再对“立定跳远”进行筛选,选择最大的 3 项

2.(2023 杭州六县九校”期中,16)小明同学为备战明年的计算机类研究生考试,收集了部分高校的计算机专业复试分数线并进行分析。小明收集到的数据如图 a 所示。

	A	B	C	D	E	F
1	院校名称	政治	外国语	业务课1	业务课2	总分
2	武汉大学	55	50	90	90	
3	南京大学	55	55	90	90	
4	复旦大学	55	55	90	90	
5	重庆大学	50	50	80	80	
6	四川大学	55	55	80	80	
7	山东大学	50	50	90	90	
8	东南大学	50	55	95	95	
9	各校平均分					

图a

各院校总分展示图



图b

请帮助小明同学对表格数据进行下列分析:

(1)若想知道这些学校的总分情况,则在 F2 单元格先输入公式_____ (要求必须使用函数),然后使用自动填充功能完成单元格 F3:F8 的计算。

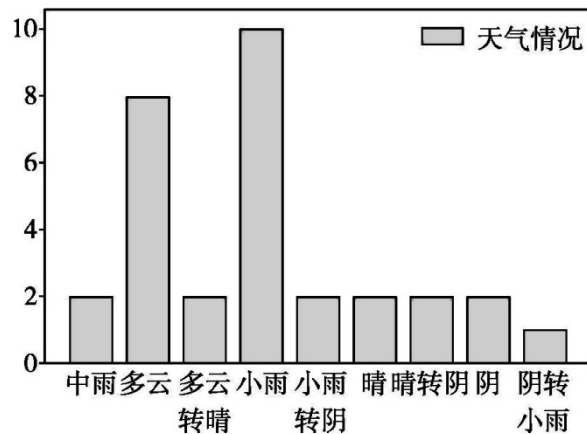
(2)现要求按照“总分”列进行降序排序,并筛选出总分在 290 分及以上的院校并显示,最后绘制各个院校总分的垂直柱形图(如图 b)。请在程序划线①②③处填入合适的代码。

```
import pandas as pd
import matplotlib.pyplot as plt
plt.rcParams['font.sans-serif']=['KaiTi','SimHei','FangSong'] #图表中文显示处理
df=pd.read_excel ("score.xlsx")
df1=df[____①____ ]
print(df1)#输出筛选数据
df=df.____②____ ("总分",ascending=False)#按照“总分”列降序排序
print(df)#输出排序好的数据
#-----创建图表代码-----
plt.title("各院校总分展示图")
plt.xlabel("院校名称")
plt.____③____ (df.院校名称,df.总分)    #各个院校总分的垂直柱形图
plt.show()
```

3.(2023 杭州 S9 联盟联考,14)小明收集了当地 2023 年 3 月份的天气情况,数据存储在文件“temp.xlsx”中,如图 a 所示。分析温差最大的日期,并生成反映各类天气情况的天数对比图如图 b 所示。

	A	B	C	D	E	F
1	日期	最高气温	最低气温	天气	风力	体感
2	2023/3/1	17	5	小雨转阴	北风4-5级	
3	2023/3/2	14	4	多云	东风3-4级	凉
4	2023-3-3	13	5	多云	微风	凉
5	2023/3/4	18	5	晴	微风	
6	2023-3-5	21	8	晴	微风	
7	2023/3/6	25	8	晴转阴	微风	
29	2023/3/28	17	10	小雨	微风	湿冷
30	2023/3/29	15	12	小雨	微风	
31	2023/3/30	13	12	小雨	微风	凉
32	2023/3/31	18	15	阴	微风	

图 a



图b

(1)在对表格进行数据整理时发现,“日期”可能存在的数据问题是_____ (单选:A.数据缺失;B.数据异常;C.逻辑错误;D.数据格式不一致)。

(2)程序代码如下所示,请在划线处填入合适的代码。

```
import matplotlib.pyplot as plt
import pandas as pd
plt.rcParams["font.sans-serif"]=["SimHei"] #设置中文字体
df=pd.read_excel("temp.xlsx")
df["温差"]=_____①_____#新增"温差"列
s=df.sort_values("温差", ascending=False, ignore_index=True)
df_max=_____②_____ #获取温差最大的日期,如并列只输出第一个日期
print(df_max)
df_t=s.groupby("天气", as_index=False).count()
df_t=df_t.rename(columns={"日期": "天数"}) #修改列名"日期"为"天数"
x=df_t["天气"]
y=df_t["天数"]
_____③_____(x, y, label="天气情况")
plt.legend()
plt.show()
```

(3)观察图 b,3 月份天数最多的天气类型是_____。

4.(2023 杭州地区重点中学期中,15)

某中学高一年级完成一次 7 选 3 意向调查,数据存储在“xk73.xlsx”中,如图 a 所示,其中 1 代表选择科目,0 代表弃选科目。

	A	B	C	D	E	F	G	H	I	J
1	学号	姓名	班级	物	化	生	政	历	地	技
2	07010101	翁*灿	—14班	1	1	0	0	0	0	1
3	07010103	潘*伟	—13班	1	1	1	0	0	0	0
4	07010103	张*恒	—14班	0	0	1	1	1	0	0
5	07010103	何*鸣	—15班	1	1	0	0	0	1	0
697	07010101	胡*艳	—2班	1	1	0	0	0	1	0
698	07010102	江*	—5班	0	0	0	1	1	1	0
699	07010102	吴*嘉	—3班	0	0	1	1	0	0	1
700	07010102	徐*	—7班	0	0	0	1	1	0	1
701	07010106	凌*瑶	—6班	1	1	0	0	0	0	1

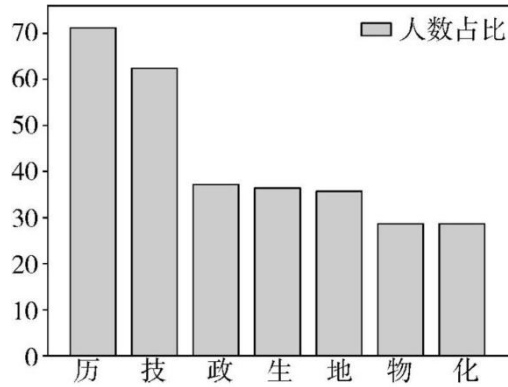
图 a

(1)使用 pandas 编程计算本次选课各门课人数占总人数的比例,请在划线处填入合适的代码。

```
import pandas as pd
import matplotlib.pyplot as plt
plt.rcParams['font.sans-serif']=['SimHei']#图表显示中文
df=pd.read_excel ("_____①_____")
a=[" "]*len(df) #存储每个学生的选课组合
dic={"物":0,"化":0,"生":0,"政":0,"历":0,"地":0,"技":0}
for i in df.index:
    for j in df.columns[3:]:
        if df.at[i,j]==1:
            _____②_____
            a[i]+=j
for i in dic.keys():
    dic[i]=round(dic[i]/len(df)*100,2)
```

(2)按照各科选考人数占比创建如图 b 所示的图表。

各科选考人数占比



图b

```
df1=pd.DataFrame({"学科":dic.keys(),"人数占比":dic.values()})
```

```
df1=_____①_____
```

```
plt.title("各科选考人数占比")
```

```
plt.bar(_____②_____, label="人数占比")
```

```
plt.legend()
```

```
plt.show()
```

观察图 b,横线处应填入的代码为:①_____;②_____(选填字母)。

A.df1.sort_values("人数占比",ascending=True)

B.df1.sort_values("人数占比",ascending=False)

C.df1.学科,df1.人数占比

D.df1.人数占比,df1.学科

(3)小李同学想查询某种七选三组合有多少人。以下程序代码可以为小李同学提供查询功能,程序运行示例如图 c 所示,请在划线处填入合适的代码。

请输入需要查询的选课组合:物化技

选择物化技组合的同学共有:192 人

图 c

```
cx=input("请输入需要查询的选课组合:")
```

```
cnt=0
```

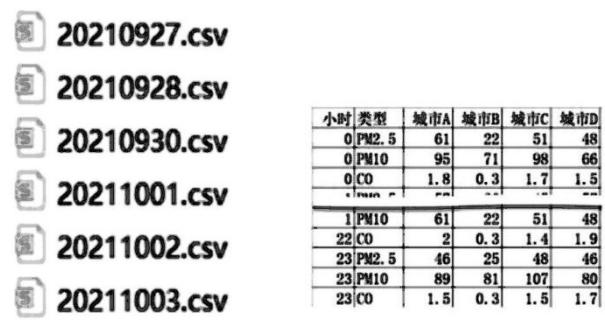
```
for i in range(len(a)):
```

```
    if _____:
```

```
        cnt=cnt+1
```

```
print(f"选择{cx}组合的同学共有: {cnt}人")
```

5.(2023 浙江 1 月选考,14,9 分)小红收集了部分城市 2021 年全年每天 PM2.5、PM10、CO 浓度数据,每天的数据分别保存在以 8 位日期字符串命名的 CSV 文件中,部分文件如图 a 所示,每个文件记录了一天 24 小时的监测数据,示例如图 b 所示。



图a

图b

为统计分析城市 A 全年各月份 PM2.5 的月平均浓度(当月的日平均浓度的平均值),编写 Python 程序。请回答下列问题:

(1)定义 pmday 函数,功能为:读取某天的 CSV 文件,返回城市 A 当天 PM2.5 的日平均浓度。函数代码如下,划线处应填入的代码为_____ (单选,填字母)。

- A.df['类型']=='PM2.5'
- B.df['类型']=='PM2.5']
- C.df[df['类型']=='PM2.5']
- D.df[df['类型']=='PM2.5']

```
import pandas as pd
def pmday(dayfile):
    df=pd.read_csv(dayfile)    #读取文件 dayfile 中的数据
    df=_____
    return df['城市 A'].mean()    #返回城市 A 当天 PM2.5 的日平均浓度
```

(2)统计城市 A 各月份 PM2.5 的月平均浓度并绘制线形图,部分 Python 程序如下,请在划线处填入合适的代码。

```
import matplotlib.pyplot as plt
def tstr(t):
    if t<10:
        retrun '0'+str(t)
    else:
```

```

    retrun str(t)

pm=[0]*12

mdays=[31,28,31,30,31,30,31,31,30,31,30,31]    #2021 年每月天数

for m in range(12):

    sm=0

    mstr=tstr(m+1)

    for d in range(①):

        dstr=tstr(d+1)

        dayfile='2021'+ mstr+ dstr+ '.csv'

        sd=pmday(dayfile)

        ②

    pm[m]=sm/mdays[m]

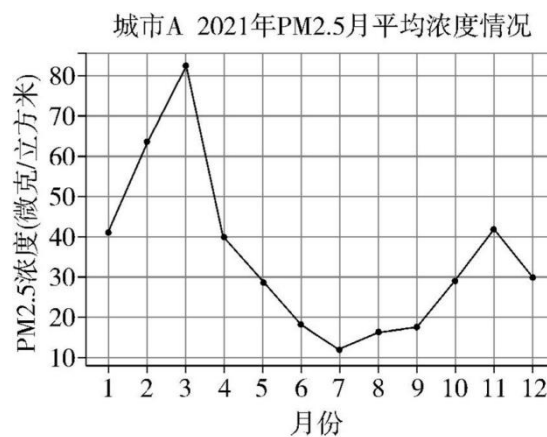
x=[1,2,3,4,5,6,7,8,9,10,11,12]

y=③

plt.plot(x,y)    #绘制线形图

#设置绘图参数,显示如图 c 所示线形图,代码略

```



图c

(3)城市 A 2021 年 PM2.5 年平均浓度为 34.6 微克/立方米。由图 c 可知,城市 A 2021 年 PM2.5 月平均浓度超过年平均浓度的月份共_____个。

考点二 大数据处理

1.(2023 宁波三锋期中,2)下列关于大数据与大数据处理的说法,不正确的是()

- A.处理大数据时,一般采用分治思想
- B.大数据的处理对象是全体数据,而不是抽样数据
- C.大数据的特点是数据量大、速度快、数据类型多、价值密度低
- D.Hadoop 是一个可运行于大规模计算机集群上的分布式系统基础架构,适用于处理实时数据

2.(2023 金华十校联考,3)下列有关大数据处理、数据可视化的说法不正确的是()

- A.静态数据是指处理时已收集完成、计算时不会发生改变的数据
- B.流数据主要是指不间断地、持续地到达的实时数据
- C.图计算是指有关大量图片的计算
- D.“各省生产总值占比情况”可以采用饼图或环形图呈现

3.有 Python 程序段如下所示:

```
import pandas as pd
import numpy as np
a=np.array([1,2,3,4]).reshape(2,2)
df=pd.DataFrame(a)
print(df.at[1,1])
```

该程序段运行后输出的结果为()

- A.4 B.3 C.2 D.1

4.小张收集了近阶段要学习的英文单词,存储为“data.txt”文件,格式如图所示。



处理“data.txt”文件中的英文单词的 Python 程序段如下所示:

```
file='data.txt'
word_c=[]
n=0
for word in open(file):
```

```
if word[0:1]=="c":  
    word_c.append(word)
```

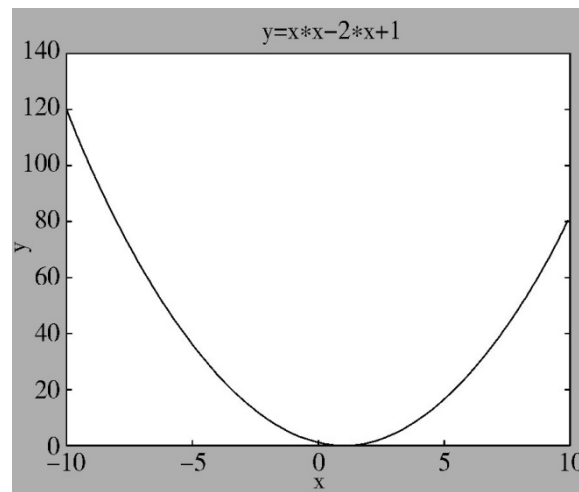
print('字母 c 开头的单词个数为:',n)

(1)划线处的代码为_____。

(2) 该程序段运行后，列表 word_c 中的数据为_____。

5.请在空格处填写正确的代码,使程序完善。

实现功能:绘制 $y=x^2-2x+1$ 的图象。



#加载 numpy 模块并取名为 np

```
import numpy as np
```

#加载 matplotlib.pyplot 模块并取名为 plt

```
import matplotlib.pyplot as plt
```

x 在-7 到 9 之间,每隔 0.1 取一个值

```
x=np.arange(-7,9,0.1)
```

```
__①__=x**2-2*x+1
```

```
plt.plot(x,__②__)
```

```
plt.title('y=x*x-2*x+1')
```

```
plt.xlabel('x')
```

```
plt.ylabel('y')
```

```
plt.__③__
```

6.在网上搜索朱自清的文章《绿》,如图所示。



(1)搜索信息并保存为 txt 文件,该过程称为_____。

(2)“绿.txt”文件的文本类型是_____。(填写字母:A.结构化数据/B.半结构化数据/C.非结构化数据)

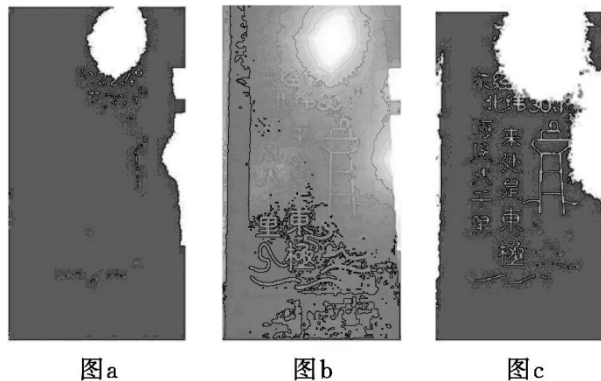
(3)制作标签云的代码如下:

```
import collections
import jieba
import wordcloud as wc
import numpy as np
from PIL import Image
wcg=wc.WordCloud(background_color="white",font_path='assets/msyh.ttf')
text=open('data/绿.txt',encoding='utf-8')
read()
seg_list=___①___
f=collections.Counter(seg_list)
wcg.fit_words(f)
wcg.to_file('output/b.png')
```

划线处①语句是调用 jieba 对象的 cut 函数对变量为 text 的文件进行分词,则该处语句为_____。

(4)得到的标签云如图所示。

小明用下列 Python 程序将图 a 处理成图 b 所示效果,发现处理后的图像不理想,他
要想将图像处理成图 c 所示效果,则可做的修改是()



```
from PIL import Image#第 0 行
import numpy as np#第 1 行
import matplotlib.pyplot as plt#第 2 行
img=np.array(Image.open ('dj.jpg').convert('L'))#第 3 行
row,cols=img.shape#第 4 行
for i in range(row):#第 5 行
    for j in range(cols):#第 6 行
        if img[i,j]>188:#第 7 行
            img[i,j]=1  #1 表示白色#第 8 行
        else:#第 9 行
            img[i,j]=0  #0 表示黑色
#第 10 行

plt.figure('dj')#第 11 行
plt.imshow(img,cmap='gray')
#第 12 行

plt.axis('off')#第 13 行
plt.show ()
```

A.将第 7 行中的数字 188 改成 138
B.将第 7 行中的数字 188 改成 250
C.将第 7 行中的>改成<
D.将第 8 行的代码与第 10 行的代码互换

3.(2022 杭州重点中学期中,13)

小萧从国家统计局网站上收集了近几年国民总收入相关数据,并使用 Excel 软件进行相关数据处理与分析。部分界面如图 a 所示,请回答下列问题:

	A	B	C	D	E	F
1	指标	2016年	2017年	2018年	2019年	2020年
2	国民总收入(亿元)	742694.1	830945.7	915243.5	983751.2	1008783
3	国内生产总值(亿元)	746395.1	832035.9	919281.1	986515.2	1015986
4	第一产业增加值(亿元)	60139.2	62099.5	64745.2	70473.6	77754.1
5	第二产业增加值(亿元)	295427.8	331580.5	364835.2	390670.6	384255.3
6	第三产业增加值(亿元)	390828.1	438355.9	489700.8	535371	553976.8
7	人均国内生产总值(元)	53783	59592	65534	70328	72000
8	国民总收入增长比例		11.88%	10.14%	7.49%	2.54%

图 a

- (1)下列关于数据整理的描述,正确的是____(单选)。
- A.某些缺失的数据可以自己随意估计一个值进行补充
- B.Excel 表格中的异常数据可以直接删除或忽略
- C.Excel 表格中的重复数据可以进行合并或删除
- D.Excel 中格式不一致的数据,一般只保留一种格式的数据,删除其他格式的数据
- (2)图 b 的图表数据类型为_____(选填:柱形图/条形图/折线图)。

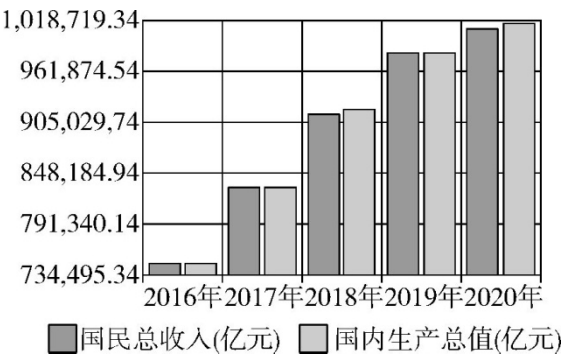


图 b

- (3)根据表格数据呈现,从 2017 年开始可以计算国民总收入增长比例,方法是在 C8 单元格输入公式_____ (计算公式:(当年国民总收入-去年国民总收入)/去年国民总收入),设置百分比格式后自动填充至 F8 单元格。
- (4)除了 Excel,可以进行数据分析的软件还有_____(多选,填数字)。
- ①Word ②SPSS ③SAS ④MATLAB
- ⑤记事本 ⑥Python ⑦录音机

4.(2022 杭州八县市区期末,16)某次测试的 Excel 文件成绩表如图 1 所示。

	A	B	C	D	E	F
1	班级	考号	姓名	信息	通用	总分
2	高二(10)班	20231001	丁*	29	35	
3	高二(7)班	20230736	黄*	42	39	
4	高二(7)班	20230701	李*	28	32	
...	...	...	...	...	...	...
197	高二(10)班	20231033	王四*	40	42	
198	高二(10)班	20231013	夏*	34	35	
199	高二(11)班	20231103	夏*	48	40	
200	高二(11)班	20231102	徐*	24	25	
201	高二(7)班	20230702	章*	32	29	
202						

图 1

(1)已知有 200 名同学参加了本次测试。小明想把全体同学的信息平均分放在 D202 单元格,那么在 D202 单元格输入的公式为_____。

(2)现在要求用 Python 增加“总分”列数据,然后求每个班总分的平均分(如图 2),最后绘制每班总分平均分的垂直柱形图(如图 3)。请在程序划线①②处选择合适的代码(填字母)。

	班级	总分
0	高二(10)班	72.75
1	高二(11)班	67.50
2	高二(7)班	67.00

图 2

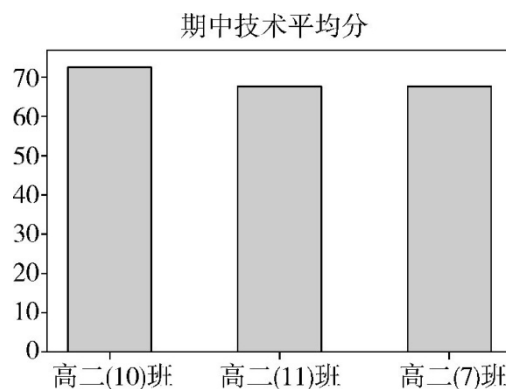


图 3

```

import pandas as pd
import matplotlib.pyplot as plt
#图表中文显示处理,代码略
df=pd.read_excel("test.xlsx")
score=[]
for i in df.values:

```

```
js=_____①_____ #①处请选择(填字母): A.df["信息"]+df["通用"]/B.i[3]+i[4]
score.append(js)
df["总分"]=score
df1=df.groupby("班级".as_index=False)["总分"].mean()
plt.title("期中技术平均分")
plt._____②_____ (df1["班级"],df1["总分"],width=0.5)#②处请选择(填字母):
A.plot/B.bar/C.scatter
```

5.(2022 丽水期末,13)小明收集了本周信息技术学科学习评价的数据,如图所示。

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	姓名	小组	选择题总分	题1	题2	题3	题4	题5	题6	题7	题8	题9	题10	题11	题12
2	杨飞田	1	26	四	2	2	2	2	2	2	2	2	2	2	2
3	邱云	1	¥18.0	2	2	2	2	0	2	2	2	2	0	2	0
4	方唯唯	1	12	2	2	0	2	2	2	0	2	0	0	0	0
5	方唯唯	1	12	2	2	0	2	2	2	0	2	0	0	0	0
6	夏向晨	2	14	2	2	0	2	2	2	0	2	0	0	0	0
28	陈哲	5	18	2	2	2	2	0	0	2	2	2	0	2	2
29	潘宏	5	16	2	2	0	2	2	2	2	2	2	0	0	0
30	方芳	5	18	0	2	0	2	2	2	2	2	2	2	0	2

(1)观察上表小明做了如下操作,其中属于数据整理的是_____ (多选,填字母)。

- A.删除重复行第五行
- B.验证并修改 D2 单元格数据
- C.通过公式计算全班平均分
- D.重新设置 C3 单元格格式

(2)为了分析每个组的平均分,设计了如下 Python 程序。

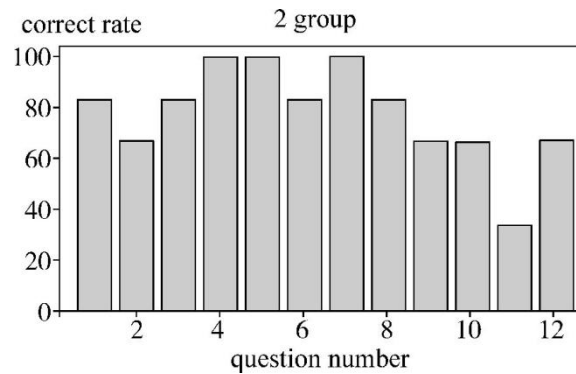
```
import pandas as pd
import matplotlib.pyplot as plt
df=pd.read_excel("成绩表.xlsx")
```

```
_____
print(df1)
```

划线处的代码应为_____ (单选,填字母)。

- A.df1=df.groupby("平均分").mean()
- B.df1=df.mean()
- C.df1=df.groupby("小组").mean()
- D.df1=df["小组"].mean()

(3)利用 Python 程序绘制各小题得分率图表,如图所示。



请在划线处填写合适的代码。

```
num=int(input("请输入小组:"))
plt.figure(figsize=(10,5))
list=[]
for i in range(12):
    s="题"+str(i+1)
    list.append(df1.at[num,s]*50)
plt.bar(range(1,13),list)
plt.title(str(num)+"_____")
plt.xlabel("question number")
plt.ylabel("correct rate")
plt.show()
```

6.(2024 届 A9 协作体返校考,14)张三同学收集了一个地区 8 月各类共享单车的骑行数据记录,每天的用户数据存储在“shared bikes. xlsx”文件中,不考虑跨天数据。数据格式如图 a 所示,请回答下列问题:

1	用户编号	App 类型	开始时间	结束时间
2	10080	青桔单车	2022/8/20 6:57	2022/8/20 7:04
3	6605	哈罗单车	2022/8/29 19:09	2022/8/29 19:31
4	10648	摩拜单车	2022/8/23 21:34	2022/8/23 21:43
5	11735	美团单车	2022/8/20 19:42	2022/8/20 19:43
41962	14826	美团单车	2022/8/20 19:42	2022/8/20 19:43
41963	6493	美团单车	2022/8/20 20:32	2022/8/20 20:52
41964	9585	青桔单车	2022/8/7 20:10	2022/8/7 20:18
41965	4906	摩拜单车	2022/8/25 23:28	2022/8/25 23:40
41966	1393	青桔单车	2022/8/23 8:55	2022/8/23 8:59

图 a

(1) cal 函数功能为:读取骑行时间的小时和分钟部分,转换为分钟格式并返回,如“2022/8/20 6:57”获取“6:57”转换为 417 ($6*60+57=417$), 代码如下。请在划线处填入合适的代码。

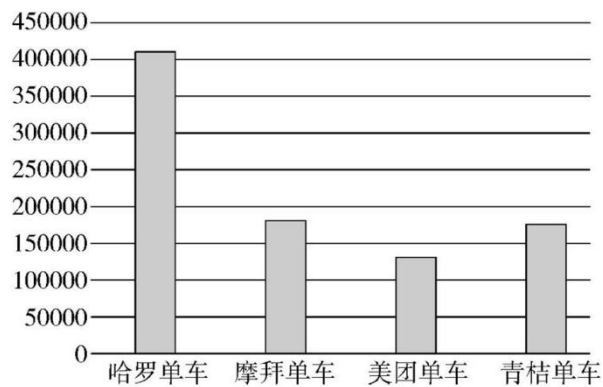
```
def cal(s):
```

```

n=len(s)
for i in range(n):
    if s[i]==" ": #如果为空格字符
        p=i
    if s[i]==":":
        q=i
t=_____+ int(s[q+1:])
return t

```

(2)统计本月各类型单车的每天平均骑行时长,并绘制柱形图,代码如下,绘制的图表如图 b 所示,请在划线处填入合适的代码。



图b

```

import pandas as pd
import matplotlib.pyplot as plt
df=pd.read_excel ("shared bikes.xlsx")
bike=["哈罗单车","摩拜单车","美团单车","青桔单车"]
sm=[0]*4
avg=[0]*4
days=31
for i in ____①____:
    t=cal(df.at[i, "结束时间"]) - cal (df. at[i, "开始时间"])
    for j in range(4):
        if df.at[i,"App 类型"]==bike[j]:
            ____②____
            break

```

```
for i in range(4):  
    avg[i]=sm[i]/ days  
plt.figure(figsize=(12, 4))  
x=bike  
y=_____③_____  
plt.bar(x, y)  
plt.show()
```

(3)统计本月各类型单车的骑行次数,加框处代码有错,可以改正为_____ (选填字母:A.max()/B.min()/C.mean()/D.count())。

```
n=df.groupby("App 类型",as_index=True).用户编号.sum()
```

7.(2024 届绍兴诊断性测试,14)学校暑期开展青春迎亚运”活动,邀请高二学生每日参加运动锻炼并进行线上打卡。每周收集一次相关数据,分别保存在相应的 xlsx 文件中,部分文件如图 a 所示。每个文件记录了一周 7 天的打卡数据,示例如图 b 所示,其中运动时长单位为分钟。



图 a

图 b

为统计分析学生锻炼情况,给出周报数据,编写 Python 程序,请回答以下问题:

(1)定义 px 函数,功能为:读取某一周的打卡数据,将其按班级进行排序操作并返回结果。

函数代码如下,将划线处代码补充完整。

```
import pandas as pd  
def px(file_week):  
    df=pd.read_excel(file_week)
```

```
df=df.sort_values(_____,ignore_index=True)
```

#按班级升序排序,参数 ignore_index=True 表示更新索引

```
return df
```

(2)统计某一周各运动项目的参与人次,并绘制柱形图,如图 c 所示,部分 Python 代码如下:

```
import matplotlib.pyplot as plt
```

```
s=input("请输入文件名: ")
```

```
df=px(s)
```

```
df1=df.groupby("运动项目".as_index=False).____①____ #统计各运动项目参与人次
```

```
df1.rename(columns={"学号":"参与人次"},inplace=True) #更改列标题
```

```
plt.bar(____②____)
```

```
plt.xlabel("运动项目")
```

```
plt.ylabel("参与人次")
```

```
plt.show()
```

划线处应填入的代码为____(单选,填字母)。

A.①count() ②df1["参与人次"], df1["运动项目"]

B.①sum() ②df1["参与人次"], df1["运动项目"]

C.①sum() ②df1.运动项目, df1.参与人次

D.①count() ②df1.运动项目, df1.参与人次

(3)统计某一周每班各学生的总运动时长后,比较得出该周每班最高的两位时长,部分 Python 程序代码如下,请在划线处填入合适的代码。

```
qp=[[0 for i in range(m)]for j in range(n)]
```

```
'''
```

定义数组 qp 记录每班各学生一周运动总时长,n 为班级数,m 为每班人数。其中 qp[0][0]~qp[0][m-1] 存储 1 班 1 号~m 号同学的每周运动总时长,依次类推,qp[n-1][0]~qp[n-1][m-1]存储 n 班 1 号~m 号同学的每周运动总时长。

```
'''
```

```
print("本周每班最高的两位时长分别为: ")
```

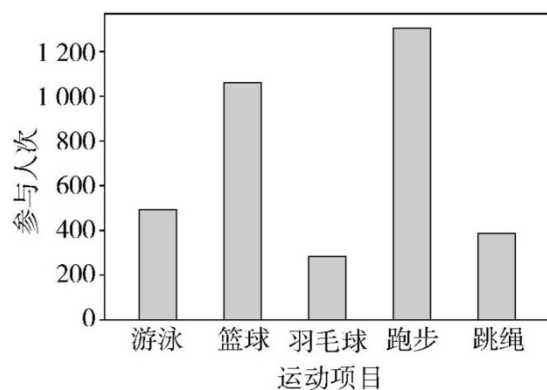
```
i=0
```

```

while i<len(df):    #df 为调用 px 函数后的返回结果
    num=df["学号"][i]
    cla=df["班级"][i]
    ①
    if i!=0 and df["班级"][i]!=df["班级"][i-1] or ②:
        cla=df["班级"][i-1]
    k1=0;k2=0
    for j in range(1,len(qp[cla-1])):
        if qp[cla-1][j]>qp[cla-1][k1]:
            ③
            k1=j
        elif qp[cla-1][j]>qp[cla-1][k2]:
            k2=j
    print(cla,"班",qp[cla-1][k1].qp[cla-1][k2])
    i=i+1

```

(4)统计某一周各运动项目的参与人次后绘制柱形图如图 c 所示,由图可知,该周参与人气最高的运动项目为_____。



图c

8.(2024届宁波模拟,13)已知某年级有6个班级,所有学生名单存储在文件“name.csv”中(如图 1),学校举行某趣味活动项目中,需要每个班抽 3 名代表参加比赛,请编写一个随机抽取程序,执行效果如图 2。请回答以下问题:

	A	B	C
1	班级	姓名	性别
2		1 林**	女
3		1 卢**	男
4		1 樊**	男
5		1 范**	女
190		6 孙**	女
191		6 孙**	男
192		6 朱**	男
193		6 汪**	女
194		6 林**	女
195		6 王**	女
196		6 陈**	女
197		6 高**	男
198		6 高**	男

图1

抽取名单为:

班级	姓名
1	张**
1	周**
1	乐**
2	魏**
2	林**
2	王**
3	乐**
3	刘**
3	曹**
4	路**
4	张**
4	丁**
5	胡**
5	闵**
5	吴**
6	王**
6	陈**
6	高**

图2

(1)加框处代码的作用是__。

(2)实现上述功能的部分 Python 程序如下,请在划线处填写合适的代码。

```
import csv
```

```
import random #数据读入
```

```
f=open("name.csv","r")
```

```
flines=csv.reader(f)
```

```
name_list=[]
```

```
m=6 #班级总数
```

```
total=18 #参赛总人数
```

```
for line in flines:
```

```
if line[0]=="班级":
```

```
continue #跳过当前循环的剩余语句,继续进行下一次循环
```

```
name_list.append(line)
```

```
_____①_____
```

```
flag=[False]*n
```

```
grade=[3]*m
```

```
i=0
```

```
while i<total:
```

```
    p=random.randint(____②____,____③____)
```

```
    bj=int(name_list[p][0])
```

```
    if flag[p]==False:
```

```
        if grade[bj-1]>0:
```

flag[p]=True

④

```
i+=1
print("抽取名单为: ")
print("班级","姓名")
for i in range(n):
    if flag[i]:
        print(name_list[i][0],name_list[i][1])
f.close()
```

9.(2023 杭州地区重点中学联考,14)小明利用“在线社团报名系统”收集了全校学生的社团报名信息,并将报名数据导出到“社团报名.xlsx”中,如图 1 所示。然后编写 Python 程序对报名数据进行处理,生成分别以班级名和社团名为文件名的 Excel 文件,以便分发给相应的社团指导老师和班主任。

	A	B	C
1	班级	姓名	社团
2	106	包大格	文学社
3	109	Jacky.Y	摄影社
4	203	来天明	书画社
...	...	...	...
211	211	李白	文学社
203	203	张瑞	书画社
209	209	李斯	足球社

图1

足球社.xlsx	1	班级	姓名	社团
文学社.xlsx	2	106	包大格	文学社
书画社.xlsx	3	106	张丽敏	文学社
摄影社.xlsx	4	109	周逸兴	文学社
摄影社.xlsx	211	211	李白	文学社
足球社.xlsx	203	203	张瑞	文学社
足球社.xlsx	212	212	袁宏道	文学社

图2

(1)在对表格进行数据整理时发现,关于“Jacky.Y” 同学的记录可能存在的数据问题是_____ (选填:A.数据缺失 B.数据异常 C.逻辑错误 D.数据格式不一致)。

(2)其中生成每个社团名单文件的过程是:先对报名数据按社团名称进行分类,并对选报同一社团的学生按班级进行升序排序,然后生成各个社团名单文件,如图 2 所示。对应的程序代码如下,请在划线处填写合适的代码。

```
import pandas as pd
def read_file(filename):
    #读入报名数据的原始文件,并将表中的数据转换成列表,代码略
def save_file(a): #保存名单到相应社团的 Excel 电子表格文件
```

```

df=pd.DataFrame(a,columns=["班级","姓名","选报社团"])
df.to_excel(____①____)+".xlsx",index=False)
a=read_file("社团报名.xlsx")
n=len(a)
#按社团名(参照拼音的字母顺序)进行升序排序,代码略
#统计各社团人数,存在列表 rs 中,rs=["滑板社",36,...],代码略
s=0
for i in range(len(rs)):
    ____②____
    left,right=s, s+num-1
    while left < right:
        imin=imax=left
        for k in range(left+1,right+1):
            if a[k][0] < a[imin][0]:
                imin=k
            elif a[k][0] > a[imax][0]:
                imax=k
        if imin !=left:
            a[imin],a[left]=a[left],a[imin]
        if imax==left:
            ____③____
        if imax !=right:
            a[imax],a[right]=a[right],a[imax]
        left=left+1; right=right-1
    ____④____
s+=num

```

题组二

1.(2022 诸暨期末,4)使用 pandas 编程处理数据 DF1,下列选项能实现行列转置操作的是()

	姓名	性别	借阅次数		0	1	2	3
0	王静	女	28	姓名	王静	张佳	李成	袁武
1	张佳	女	56	性别	女	女	男	男
2	李成	男	37	借阅次数	28	56	37	68
3	袁武	男	68					

- A.DF1.T
 B.DF1.columns
- C.DF1.values
 D.DF1.index

2.(2022 宁波奉化期末,9) 有如下 Python 程序段:

```

import pandas as pd
s=pd.Series(range(5,11,3))
s[1]=15
print(s)
    
```

该程序执行后,输出的结果是()

1 15 2 8 dtype: int64	0 5 1 15 dtype: int64	1 15 2 8 3 11 dtype: int64	0 5 1 15 2 11 dtype: int64
--	--	---	---

- A.
 B.
 C.
 D.

3.(2022 宁波咸祥中学期中,16)小明收集了某超市饮料销售情况相关数据,并使用 Excel 软件进行处理,如图 a 所示。

	A	B	C	D	E	F	G	H	I
1	超市饮料类销售情况统计表								
2	类别	饮料名称	数量	单位	进价 (元)	售价 (元)	销售额 (元)	毛利润 (元)	占销售总 额百分比
3	茶	康师傅绿茶	56	瓶	2.20	2.75	154.00	30.80	5.65%
4		康师傅冰红茶	19	瓶	2.20	2.80	53.20	11.40	1.95%
5	橙汁	统一鲜橙多	64	瓶	2.10	2.80	179.20	44.80	6.57%
6		芬达橙汁	76	瓶	2.20	2.80	212.80	45.60	7.80%
7		康师傅鲜橙多	66	瓶	2.00	2.80	184.80	52.80	6.78%
8	可乐	可口可乐	75	瓶	5.80	6.45	483.75	48.75	17.74%
9		百事可乐	50	瓶	5.50	6.50	325.00	50.00	11.92%
10		罐装可口可乐	76	罐	2.10	2.50	190.00	30.40	6.97%
11		罐装百事可乐	73	罐	2.10	2.45	178.85	25.55	6.56%
12	凉茶	王老吉	86	罐	2.10	3.00	258.00	77.40	9.46%
13		和其正	65	罐	2.10	3.00	195.00	58.50	7.15%
14		盒装王老吉	49	盒	1.00	2.00	98.00	49.00	3.59%
15	雪碧	罐装雪碧	84	罐	2.00	2.55	214.20	46.20	7.86%
16	销售总额(元):							2726.8	

图 a

请回答下列问题:

(1) 区域 I3:I15 的数据是通过公式计算得到的。在 I3 单元格输入公式后,用自动填充功能完成 I4:I15 的计算,则 I3 单元格中的公式是_____。(占销售总额百分比=销售额(元)/销售总额(元))

(2) 若因误操作将 I16 单元格删除,则 I3 单元格会显示_____(选填字母:A.#VALUE B.#DIV/0! C.##### D.#REF!)。

(3) 要将类别为“可乐”的饮料以“毛利润(元)”为主要关键字降序排序,则选择排序的区域为_____。

(4) 小王根据图 a 中的数据,制作了一张反映橙汁类饮料销售额和毛利润对比的图表,如图 b 所示,则建立该图表的数据区域是_____。将 E8 单元格数值修改成 6.00,是否会影响该图表?_____(选填:是/否)

橙汁类饮料销售额和毛利润对比图

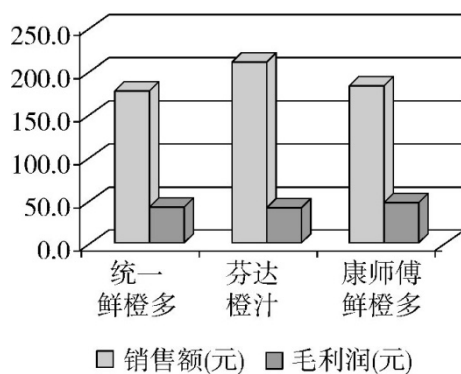


图 b

4.(2022 台州启超中学期中,16)APP 活跃人数数据存储在“app.xlsx”文件中,如图所示,现要编程统计结果,请回答下列问题:

	A	B	C	D	E
1	APP名称	应用领域	10月人数	11月人数	12月人数
2	360手机卫士	安全管理	13 855.8	14 422.3	16 017.2
3	腾讯手机管家	安全管理	17 992.5	18 135.2	19 688.5
4	微信	即时通讯	74 257	75 101.6	75 616.5
5	QQ	即时通讯	53 931.5	54 913.7	55 216.7
6	QQ浏览器	浏览器	20 759.1	20 540.3	22 438.6
7	手机百度	搜索	18 401.2	18 571.3	19 960.2
8	QQ音乐	移动音乐	13 707.9	13 709.1	13 691.5
9	酷狗音乐	移动音乐	19 407.5	19 237.8	20 847.05
10	酷我音乐	移动音乐	4 935.37	4 864.13	5 480.75
11	支付宝	支付	29 897	31 408.9	33 146.8
12	京东	综合电商	6 441.72	7 596.09	6 854.08
13	淘宝	综合电商	27 153.9	27 737.6	29 807.4
14	爱奇艺视频	综合视频	20 319.4	20 543.8	22 047.9
15	芒果tv	综合视频	5 283.03	5 165.8	5 108.08
16	腾讯视频	综合视频	19 833.5	20 412.6	21 332.2
17	优酷视频	综合视频	12 576.4	12 265.5	13 840.4

若要把“app.xlsx”第 1 张工作表中的信息导入到 book1 对象中,并进行统计,实现这个功能的 Python 代码如下,在程序划线处填入合适的代码。

```
import pandas as pd

book1=_____ (1) _____ #读取 app.xlsx 文件数据,并存储在 book1 对象中

book1_sum=_____ (2) _____

#计算 10 月人数之和

book1_aver=_____ (3) _____

#计算 11 月人数平均值

book1_g=_____ (4) _____

#按应用领域分组统计

book1_sort=_____ (5) _____

#按 11 月人数值,降序排序

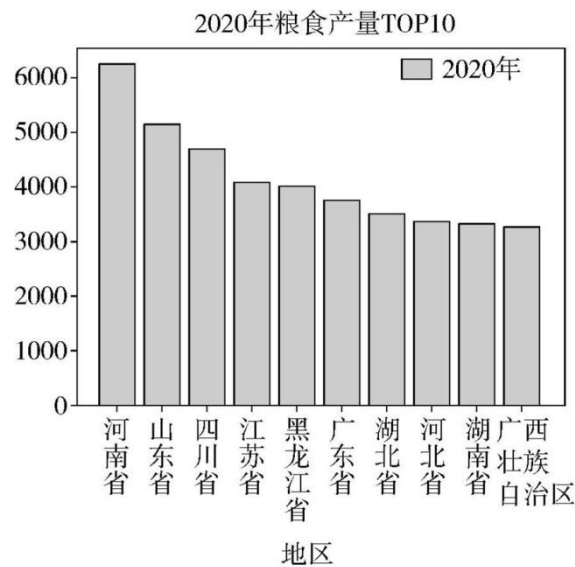
print("10 月人数之和:",book1_sum)

print("11 月人数平均值:",book1_aver)
```

5.(2022 宁波奉化期末,14)小孙收集了 2016 年到 2020 年的各地区粮食生产总量并存储在“lscl.xlsx”文件中,如图 a 所示,现使用 Python 对其进行数据处理,并实现数据可视化,绘制的图表如图 b 所示。

	A	B	C	D	E	F
1	地区	2020年	2019年	2018年	2017年	2016年
2	北京市	107.57	102.33	114.75	129.83	145.2
3	天津市	228.76	202.91	197.21	183.17	181.89
4	河北省	3413.34	3114.86	3085.86	2890.6	2772.86
5	山西省	1075.94	936.76	894.95	861.89	824.42
6	内蒙古自治区	1699.01	1606.34	1512.5	1434.73	1477.56
7						
29	陕西省	2807.07	2445.83	2244.96	2095.29	1997.81
30	甘肃省	1423.85	1306.41	1166.1	1068.61	985.73
31	青海省	188.6	181.25	169.24	162.38	155.52
32	宁夏回族自治区	397.91	330.78	344.63	308.96	299.71
33	新疆维吾尔自治区	2936.33	2616.3	2541.16	2313.24	2201.74

图a



图b

实现如上功能的代码如下,请回答以下问题。

```
import pandas as pd
import matplotlib.pyplot as plt
plt.rc('font', **{'family': 'SimHei'})
#设置中文字体
df=pd.read_excel("lscl.xlsx")
df._____①_____("2020 年",ascending=False,inplace=True)
df1=df.head(10)
x=_____②_____
y=df1["2020 年"]
plt.figure(figsize=(8,6))
plt._____③_____("2020 年粮食产量 TOP10")
plt.bar(x,y,label="2020 年")
plt.xlabel("地区")
```

plt.legend()

④

(1)请在划线处填入合适的代码语句,以实现以上功能。
(2)代码语句“plt.bar(x,y,label="2020 年")”的功能为绘制如图 b 所示的图像,能够实现相同功能的语句是_____ (多选题,少选得 1 分,多选不得分)。

- A.df.head(10).plot("地区","2020 年")
- B.df[10:].plot("地区","2020 年",kind="bar")
- C=plt.plot(x,y,label="2020 年",kind="bar")
- D=plt.bar("地区","2020 年",label="2020 年")
- E=df1.plot("地区","2020 年",kind="bar")

6.(2022 杭州场口中学、桐庐富春中学检测,16)在月考之后,学校教务员拿到了高一年级月考的基础成绩并用 Excel 软件进行数据处理,如图 1 所示。请回答下列问题:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	考号	班级	姓名	语文	数学	英语	政治	历史	地理	物理	化学	生物	总分
2	411302	→ (13) 班	李金洲	116	106	88.5	68	80	79	66	68	67	738.5
3	411335	→ (13) 班	徐闻	109	60	122.5	63	80	68	29	35	61	627.5
4	410206	→ (2) 班	毛雨雪	108	67	93.5	67	64	65	33	42	57	596.5
5	411209	→ (12) 班	羊一	108	87	123.5	61	84	53	53	58	65	692.5
6	410535	→ (5) 班	倪玲意	107	84	49	55	70	57	37	50	61	570
7	411210	→ (12) 班	刘子晨	105	81	105.5	56	76	63	52	53	78	669.5
641	410142	→ (1) 班	袁天昂	62	86	46	36	0	34	43	50	41	398
642	410342	→ (3) 班	陆峻泽	60	41	40.5	56	48	27	33	24	18	347.5
643	411334	→ (13) 班	王坤	59	81	110.5	52	61	58	61	46	62	590.5
644	411036	→ (10) 班	包雨婷	54	82	68	39	61	52	37	49	46	488
645	410544	→ (5) 班	李佳乐	53	23	25	26	48	42	19	29	30	295

图 1

(1)区域 M2:M645 的数据是通过在 M2 单元格输入公式并自动填充得到的,则 M645 单元格中的公式是_____。
(2)为了分析各班级的数学平均分,设计了如下 Python 程序,利用其绘制各班数学平均分图表,如图 2 所示:

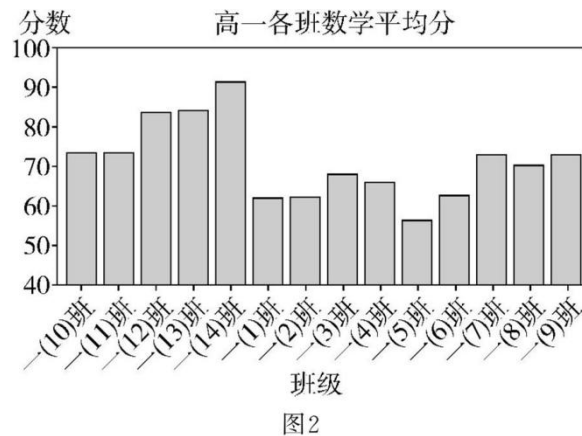


图2

结合上图的效果,请在程序划线处填写合适的代码。

```
import matplotlib.pyplot as plt
import pandas as pd
df=pd.read_excel('高一月考成绩.xlsx')
df1=df.groupby("班级"). _____①_____
plt.bar(df1.index,_____②_____)
plt.title("高一各班数学平均分")
plt.ylim(40,100)
plt.xlabel(_____③_____)
plt.ylabel("分数")
plt.show()
```

7.(2023 三月百校联考,14)李明收集了梅西 2004 年至 2022 年俱乐部比赛数据,保存在梅西俱乐部详细比赛数据.xlsx”文件中,部分数据如图 a 所示,现在利用 pandas 模块处理数据。

	A	B	C	D	E	F	G	H	I	J	K
1	序号	时间	赛事	主队	比分	客队	阵容	上场时间	进球	助攻	得牌
2	1	2004-10-16	西甲	西班牙人	0-1	巴塞罗那	替补	7	0	0	0
3	2	2004-10-24	西甲	巴塞罗那	3-0	奥萨苏纳	替补	18	0	0	0
4	3	2004-10-27	国王杯	格拉马内	1-0	巴塞罗那	首发	72	0	0	0
5	4	2004-12-04	西甲	巴塞罗那	4-0	马拉加	替补	8	0	0	0
6	5	2004-12-07	欧冠	顿涅茨克	2-0	巴塞罗那	首发	90	0	0	0
7	6	2004-12-11	西甲	阿尔巴塞	1-2	巴塞罗那	替补	1	0	0	0
8	7	2004-12-21	西甲	巴塞罗那	2-1	莱万特	替补	33	0	0	0
9	8	2005-04-17	西甲	巴塞罗那	2-0	赫塔菲	替补	5	0	0	0
10	9	2005-05-01	西甲	巴塞罗那	2-0	阿尔巴塞	替补	3	1	0	0

图 a

(1)梅西 2004 年 10 月—2021 年 7 月效力于巴塞罗那俱乐部,2021 年 8 月转会至巴黎圣日耳曼俱乐部,现在李明想知道梅西每个赛季的胜率,实现上述功能的 Python 程序如下,请在划线处填入合适的代码。

```
import pandas as pd
```

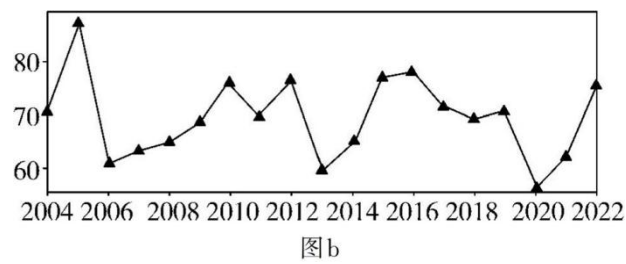
```

import matplotlib.pyplot as plt
df=pd.read_excel("梅西俱乐部详细比赛数据.xlsx")
df['年']=pd.to_datetime(df["时间"]).dt.strftime('%Y').astype(int)
df['月']=pd.to_datetime(df["时间"]).dt.strftime('%m').astype(int)
df['胜负情况']=" "
for i in range(len(df)):
    f=True;z=0;k=0
    for c in df["比分"][i]:
        if ①:
            if f:
                z=z*10+int(c)
            else:
                k=k*10+int(c)
        else:
            f=not f
    jlb="巴塞罗那"
    if df["年"][i]*100+df["月"][i]>=202108:
        ②
    if (df["主队"][i]==jlb and z>k) or (df["客队"][i]==jlb and z<k):
        df['胜负情况'][i]="胜"
    elif z==k:
        df['胜负情况'][i]="平"
    else:
        df['胜负情况'][i]="负"
g=df.groupby(df["年"],as_index=True).count()
③
g1=df1.groupby(df1["年"],as_index=True).count()
g["胜率"]=g1["胜负情况"]/g["胜负情况"]*100

```

(2)图 b 为 2004—2022 年梅西俱乐部比赛胜率统计图。

2004-2022 年梅西俱乐部比赛胜率统计



```
plt.plot(, marker='^')
plt.title("2004-2022 年梅西俱乐部比赛胜率统计")
plt.show()
```

- 方框处代码为_____ (多选:填字母)。
- A.g["年"],g["胜率"]
 - B.g["年"],g.胜率
 - C.g.index,g["胜率"]
 - D.g.index,g.胜率

8.(2022 湖州三贤联盟期中,14)小张同学为了更好地了解冬奥会,从网上收集了历届冬奥会各个项目比赛信息,收集到的部分数据如图 1 所示。

序号	姓名	性别	国家	届次	运动项目	奖牌
1	Christine Jacoba Aaftink	女	荷兰	16届	速度滑冰	0
2	Christine Jacoba Aaftink	女	荷兰	16届	速度滑冰	0
3	Christine Jacoba Aaftink	女	荷兰	17届	速度滑冰	0
4	Christine Jacoba Aaftink	女	荷兰	17届	速度滑冰	0
5	Christine Jacoba Aaftink	女	荷兰	18届	速度滑冰	0
6	Christine Jacoba Aaftink	女	荷兰	18届	速度滑冰	0
7	An Yulong	男	中国	19届	短道速滑	S
8	An Yulong	男	中国	19届	短道速滑	0
9	An Yulong	男	中国	19届	短道速滑	8
10	An Yulong	男	中国	20届	短道速滑	B

图 1

中国历届冬奥会奖牌趋势图

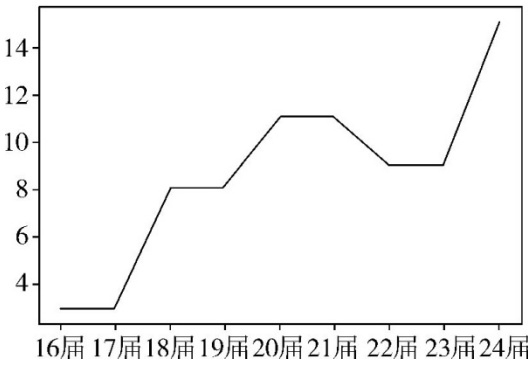


图 2

为分析数据,小张编写了如下程序:

```
import pandas as pd

import matplotlib.pyplot as plt

from matplotlib.pyplot import MultipleLocator

plt.rcParams['font.sans-serif']=['SimHei'] #正常显示中文标签

df=pd.read_csv("dongao.csv")

#删除所有未获得奖牌的记录,并将奖牌列中的"G"修改为"金牌","S"修改为"银牌"

","B"修改为"铜牌"

jp={'G': '金牌','S': '银牌','B': '铜牌'}

for i in df.index:

    if ①:

        df=df.drop(i)

    else:

        df.at[i,'奖牌']=jp[df.at[i,'奖牌']]

#对输入国家每届的奖牌数进行统计,并制作相应图表,如图 2 所示

nt=input("请输入国家名称:")

df1=df[df['国家']==nt]



df3=pd.DataFrame(df2) #将分组后数据生成新的二维结构,索引为“届次”,列标题

为“奖牌”

x=df3.index

y=②

plt.title(nt+"历届冬奥会奖牌趋势图")

plt.③(x,y)

plt.show()
```

(1)在划线处填上合适的代码。

(2)为了能显示某国历届冬奥会奖牌变化,需在加框处添加的语句为_____ (多选)。

A.df2=df1.groupby('奖牌')

```
df2=df1.届次.count()
B.df2=df1.groupby('届次')
df2=df2['奖牌'].count()
C.df2=df1.groupby('奖牌')['届次'].count()
D.df2=df1.groupby('届次').奖牌.count()
```

9.(2023 十校联盟联考,14)小明从网上下载了豆瓣图书 1900—2017 年间出版的图书数据,存储在 Excel 文件中,如图 a 所示,数据表已按出版年份升序排好,包含书名、作者、出版社、出版年份、价格、评分以及评论数量。他要编写一个 Python 程序快速对图书数据进行分析。

	A	B	C	D	E	F	G	H
1	书名	作者	出版社	出版年份	价格	评分	评论数量	
2	岳代斯玛先生的午后	〔法〕杜拉斯	上海译文出版社	1900	16	7.4	1262	
3	巴黎线	文达姆	商务印书馆	1900	21	7.4		
4	巴城线	文达姆	商务印书馆	1900	21		90	
5	版通纪程鲁迅藏中国现代木刻	上海鲁迅纪念馆	江苏古籍出版社	1900	988	0	0	
6	黑湖人时代	皮力著	江苏美术出版社	1900	15	7.2	69	
7	成功人士的七个习惯	史蒂芬·柯维	中信出版社	1900	28	8	24	
8	长堤	马智康	海南文艺出版社	2017	29.8	0	0	
56165	长堤	〔美〕阿德里·克拉克	北京联合出版公司·磨房出版	2017	42	0	0	
56166	长夜难明	紫金陈	湖南文化/云南人民出版社	2017	42	0	0	
56167	雪中花园	张筱亮	中信出版社	2017	68	7.6	12	
56168	中国日记	〔美〕斯蒂芬·斯	浙江人民美术出版社	2017	128	None	0	
56169	中国通史	张应麟	台海出版社	2017	108	9.9	47	
56170	神游院植物。安守四季阳光	熊秉真	时代华文书局	2017	39.8	0	0	
56171	追风筝的回忆	韩九黎	北京联合出版公司/联合读创	2017	38	0	0	
56172								

图 a

- (1)为了求评论数量累计最高的作者及其出版的图书平均评分,小明需要对图 a 所示的表中数据进行整理,则下列说法正确的是_____ (多选,填字母)。
- A.第 3 行和第 4 行数据重复,删除其中一行即可
 - B.通过检测发现 F56168 单元格的数据存在错误,应进行修正
 - C.删除“出版社”和“出版年份”两列数据,不影响分析结果
 - D.“评分”及“评论数量”为 0 的数据没有任何价值,可以直接删除
- (2)小明利用整理好的数据,编写并运行程序,结果如图 b 所示。

	作者	评论数量
0	[日]村上春树	679101.0
1	[日]东野主吾	664106.0
2	韩寒	623116.0
3	郭敬明	620571.0
4	[英]J.K.罗琳	462476.0

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。

如要下载或阅读全文，请访问：

<https://d.book118.com/518030027142006050>